



International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 2, January 2013)

National conference on Machine Intelligence Research and Advancement (NCMIRA, 12), INDIA.

Web Page Grader of the Website

GEETA R.B.¹, SHOBHA R.B.², Prof. SHASHIKUMAR G.TOTAD³, Dr. PRASAD REDDY PVGD⁴

¹Department of Information Technology, GMRIT, RAJAM, AP, India

²Department of Electronics and Communication, BEC, Bagalkot, Karnataka, India

³Department of Computer Science, GMRIT, RAJAM, AP, India

⁴Department of CS & SE, Andhra University, Vizag, AP, India

geetatotad@yahoo.co.in¹, srb_bec@yahoo.com², skumartotad@yahoo.com³, prasadreddy.vizag@gmail.com⁴

Abstract

The Web Page Grader guides website designers/administrators to evaluate the web page by collecting various statistics from the Log file of the server and knowing the structure of the website. Every request of user for web information will be stored in the Log File. Along with statistics collected the structure of the web site plays a vital role. If the popular page is located deep into the web site users may not get quick response. It consumes server's time, client's time and more power. It takes more bandwidth. Along with this, too many transactions will be entered into the Log File which takes server's more memory space. To overcome these problems, a novel approach is proposed which helps in providing the decision about popularity of the web page based on Parameter grading. The parameters can be classified into subjective and objective. The subjective parameters include level of web page, number of backward links to the web page. The objective parameters include frequency and utility.

Keywords- Website, Utility, WP Grade, Log File

I. INTRODUCTION

The World Wide Web serves as an enormous, extensively distributed, global information repository. Web mining can be broadly classified into three categories i.e is joint family of web structure mining, Web Usage Mining and Web Content Mining [1]. Web mining is a popular technique for analyzing visitor activities in e-service systems [2]. A Website is a collection of web pages. Whenever user requests for a web page using Hypertext transfer protocol (HTTP) transaction, user request is processed by server, its entry is made in the Log file containing various fields such as IP address, time at which request is made, which Uniform Resource Locator(URL) is requested, status code. The website is kept under observation for some time. It is observed offline mainly to collect the statistics of various parameters such as what is time spent on each page, how many users visited a particular page, how many failure status codes are there etc...

II. RELATED WORK

The web site's evaluation can be accomplished by user's visits evaluation, in order to assist in this, many commercial systems provide statistics about the most visited files and pages.

A significant recent work is that the pages with more hit count can be given the preference to be brought closer to the home page provided web site owner/designer agrees [5]. Miller and others [3] suggested that the structure of linked pages has decisive impact on the usability. However in [4], it is revealed that, the number of hits per page, calculated from log file processing, is an unreliable indicator of page popularity. The number of pages at the same level, number of forward links and the number of backward links to a particular web page reflect visitor's behavior during navigation to a specific website [6]. Thus, a robust metric is proposed, which takes into account structural information, and its utility from the log file of the server is considered, certain pages are reorganized leading to the overall improvement to site access.

III. PROPOSED WORK

Consider a web site containing seven web pages as shown in Fig. 1. The website is kept under observation for specific period of threshold. The various statistics are collected from the Log File of the server such as how many pages are accessed by various users, what is the time spent on each web page etc. The frequency and utility collected from the Log file of server is shown in Fig.1. The frequency & utility of web page A is 20 & 35 respectively.



International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 2, January 2013)

National conference on Machine Intelligence Research and Advancement (NCMIRA, 12), INDIA.

Total frequency (Tot_Fr) of the web site is 213 & Total Utility (Tot_U) is 153. Consider the frequency & utility obtained from the Log File of the server. This data does not give any useful information. The collected data may create confusion in mind. These are raw data or ungrouped data.

Web page grading is calculated considering 4 parameters frequency, utility, level and number of backward links for a particular web page. The web page with more backward links indicates this web page is of more importance.

The weightages for subjective and objective parameters of the web page of the given website are assumed as shown in Table I. PrG_i stands for Parameter Grade of web page for i^{th} parameter. The four parameters assumed are frequency, Utility, Backlinks and Level of Web page.

Table I

Subjective and Objective Parameters Weightages of web page of the website

Sl. No	Parameters	Weightages
1	Frequency	200
2	Utility	300
3	Backlinks	300
4	Level of Web Page	200
Total		1000

Equation (1) is used to assign a grade to the web pages of the website using four parameters frequency, utility, Level, number of backlinks, Total frequency, Total Utility, Level of a page $L(W_i)$ and total number of levels in the website Tot_L.

$$F = (Fr(W_i) / Tot_Fr) * PrG_1$$

$$U = (Util(W_i) * PrG_2) / Tot_U$$

$$FBL = (BL(W_i) / FL(W_i)) * PrG_3$$

$$L = (L(W_i) / Tot_L) * PrG_4$$

$$WPG = (F + U + FBL + L) / 1000 \quad (1)$$

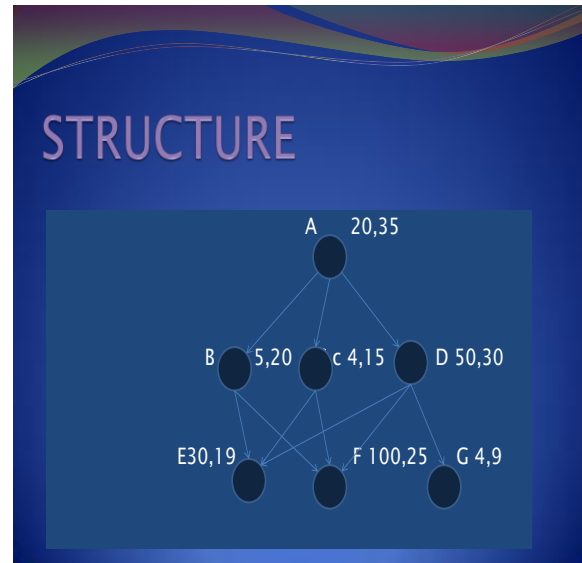


Fig 1. Structure of the website

The grade of web pages of the website calculated using “equation (1)” is as follows.

Table II

WPG values for each web page

Sl.No	Web page Name	WPG
1	A	0.07
2	B	0.33
3	C	0.32
4	D	0.34
5	E	0.66
6	F	0.74
7	G	0.22

The web pages F, E and D are popular pages and these pages can be brought near to the home page as shown in Fig. 2. Fig.2 shows the structure of the web site after reorganization.



International Journal of Emerging Technology and Advanced Engineering

Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 2, January 2013)

National conference on Machine Intelligence Research and Advancement (NCMIRA, 12), INDIA.

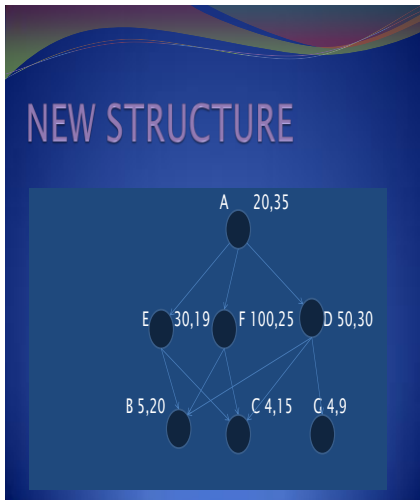


Fig. 2 Structure of Reorganized website

IV. CONCLUSION

In this paper we propose a new metric which grades the web page based on subjective and objective values. If the most popular page is very near to the leaf node or away from the parent node, users won't get quick response. To reach leaf node user has to travel from root to leaf, so many times communications happen between client and server, so many log entries will be made into server. The bandwidth utilized will be more.

Too many control messages must be generated between server and client, which involve client's time and server's time. Delay incurred will be more. To overcome these problems, popular pages can be moved near to home page. We invite extensions and vital improvements in the metric.

REFERENCES

- [1] B Cooley, Mobasher B and Srivastava J, "Web Mining: Information and pattern discovery on the World Wide Web[C]", Proc of Int'l Conf on Tools with Artificial Intelligence (ICTAI'97), 558-567 (1997).
- [2] Li, L., Zhang, C. and Zhang Cooperative Strategy Web-Based Data Cleaning, Applied Artificial Intelligence, Vol. 17, Nos. 5-6, 443-460 (2003).
- [3] Miller, C.S. and Remington, R. W. Implications for information Architecture, Human Computer Interaction, Journal IEEE Web Intelligence: Vol. 19, No.3, 225-271 (2004).
- [4] Garofalakis, Web Site Optimization Using Page Popularity, IEEE Internet Computing, Vol 3, No. 4, 22-29 (1999).
- [5] Geeta.R.B, Shashikumar G. Totad & Prasad Reddy PVGD, Amalgamation of Web Usage Mining and Web Structure Mining, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, 279-281 (2009).
- [6] Geeta R.B, Shashikumar G. Totad & Prasad Reddy PVGD, Optimizing User's Access To Web Pages, RJoosJA, Transactions on World Wide Web- Spring, Vol 8, No. 1, 61-66 (2008).
- [7] Jing Wang, Ying Liu, Lin Zhou, Yong Shi, and Xingquan Zhu, Pushing Frequency Constraint to Utility Mining Model, ICCS Springer-Verlag Berlin Heidelberg, LNCS 4489, 685-692 (2007).